Please replace the paragraph on page 2, lines 1-7, with the following paragraph.

The resulting molecules, while serving as lead compounds, often have unpredictable effects when employed in clinical trials. In addition, it has been observed that existing drugs with known clinical efficacy far often fail to achieve beneficial results when given to particular patients, or particular populations, such as ethnic groups, of patients. Genetic stratification of a population can be the difference between drug failure and drug approval.

A1

Please replace the paragraphs beginning on page 3, line 23, through page 4, line 11, with the following paragraphs.

Genetic polymorphisms arise, for example, as a result of gene sequence differences or as a result of post-translational modifications, including glycosylation. Hence genetic polymorphisms are manifested as gene products and proteins having variant structures. The variant structures result in differences in biological responses among the originating organisms. These differences in response, include, but are not limited to, differences among patient responses to a particular drug, effective dosage differences, and side effects. With respect to infectious organisms, some polymorphisms may arise that convey resistance or susceptibility to particular drug therapies by altering the drug target structure.

A2

Structural changes that arise as a result of genetic polymorphisms, are not of unlimited variety, since 3-D structure impacts upon function. A knowledge of the repertoire of the fine differences among generally similar 3-D structures of particular proteins will permit design of drugs that bind to the most polymorphisms, drugs that induce the fewest side-effects, and drugs that are more effective against infectious agents. Knowledge of these structures ultimately will permit patient-specific or subpopulation-specific, such as ethnic groups or age group, design or selection of drugs.

Please replace the paragraph beginning on page 4, line 23, through page 5, line 5, with the following paragraph.

After the protein structural variant models are generated, selected model structures can be analyzed to determine common structural features that are conserved throughout the selected models. The conserved structural features can serve as scaffolds or pharmacophore models into which potential drugs or modified drugs are docked. For example, the selected model structures may represent the structural variants resulting from the most commonly occurring genetic polymorphisms or from genetic polymorphisms found in a specific patient subpopulation. Alternatively, the models may be selected based on clinical information; for example, the structural variants may be derived based on patients receiving a specific treatment regimen or exhibiting a particular clinical response to a given drug, or on the duration of a particular drug treatment, or a particular age group or ethnic or racial group, sex or other subpopulation.

A3

Please replace the paragraph beginning on page 6, line 27, through page 7, line 7, with the following paragraph.

Molecular structure databases containing protein structural variant models produced by the methods are also provided. The databases may also contain biological or clinical data associated with the structural variants. The databases can be interfaced to a molecular graphics package for visualization and analysis of the 3-D molecular structural models. In particular, databases containing the 3-D structures of polymorphic variants of selected target genes, particularly pharmaceutically significant genes, such as proteases and polymerases, including reverse transcriptases, and receptors, such as cell surface receptors, are provided. The databases may be stored and provided on any suitable

A4

A4 cont'd

medium, including, but are not limited to, floppy disks, hard drives, CD-ROMS and DVDs.

Please replace the paragraphs beginning on page 8, line 22, through page 9, line 14, with the following paragraphs.

A polymorphic marker or site is the locus at which divergence occurs. Such site may be as small as one base pair (a SNP). Polymorphic markers include, but are not limited to, restriction fragment length polymorphisms, variable number of tandem repeats (VNTR's), hypervariable regions, minisatellites, dinucleotide repeats, trinucleotide repeats, tetranucleotide repeats and other repeating patterns, simple sequence repeats and insertional elements, such as Alu. Polymorphic forms also are manifested as different mendelian alleles for a gene. Polymorphisms may be observed by differences in proteins,

A5

protein modifications, RNA expression modification, DNA and RNA methylation, regulatory factors that alter gene expression and DNA replication, and any other manifestation of alterations in genomic nucleic acid or organelle nucleic acids.

As used herein, structural variant proteins refer to a variety of 3-D molecular structures or models thereof that result from the polymorphisms. These variants typically arise from transcription and translation of genes containing genetic polymorphisms.

As used herein, binding interactions refer to atomic or physical interactions between molecules including, but not limited to, binding free energy, hydrophobic interactions, electrostatic interactions, steric interactions and other interactions that are commonly considered by those of skill in the art to determine the affinity of one molecule to bind to another. Favorable binding interactions refer to binding interactions that promote physical or chemical associations between molecules.

Please replace the paragraphs on page 9, lines 19-27, with the following paragraphs.

As used herein, structure-based drug design refers to computer-based methods in which 3-D coordinates for molecular structures are used to identify potential drugs that can interact with a biological receptor. Examples of such methods include, but are not limited to, searching of small molecule libraries or databases, conformational searching of a ligand within an active site to identify biologically active conformations or computational docking methods.

A6

As used herein, pharmacogenomics refers to the study of the variability of patient responses to drugs due to inherent genetic differences.

Please replace the paragraph on page 11, lines 1-8, with the following paragraph.

A. Structure generation and analyses

As noted, patients exhibit variable responses to drugs. For some patients a drug may be very beneficial and achieve a desired response; whereas for other patients, with the same disorder, the same drug will have little or no effect. It is known that individuals as well as groups of individuals exhibit a variety of genetic polymorphisms. As described herein, the presence or absence of such polymorphism can be correlated with the variability of patient responses to drugs.

A7

Please replace the paragraph beginning on page 11, line 25, through page 12, line 3, with the following paragraph.

It is shown herein that it is advantageous to utilize 3-D molecular structures in drug design rather than to consider sequences alone. For example, most drugs target proteins, and disease, drug action and toxicity are all manifested at the protein level. Although the nucleotide sequences of genetic polymorphisms might appear to be quite different, the resulting protein targets may have similar shapes and, therefore, the protein biological function might be the same. Conversely, although genetic polymorphism sequences might appear similar, the resulting proteins may have critical differences in their 3-D structures that greatly affect biological activity.

A8

Please replace the paragraph beginning on page 12, line 20, through page 13, line 2, with the following paragraph.

1. Generating 3-D protein structural variant models

The first step in the methods provided herein is to obtain patient samples of a gene that exhibits genetic polymorphisms or of a therapeutic target protein derived therefrom. Starting with gene sequences that include single or multiple nucleotide polymorphisms, the amino acid sequences of the translated proteins can be determined. Alternatively, patient samples of the target protein can be A9 obtained and sequenced directly. Multiple sequence analyses can be performed to determine the exact amino acid variations or mutations resulting from the genetic polymorphisms. Numerous methods for identifying genes that encode polymorphisms are known, numerous polymorphisms have been identified and mapped, and databases of such polymorphisms are publicly available.

Please replace the paragraphs beginning on page 16, line 4, through page 17, line 30, with the following paragraphs.

A preferred method for generating and refining the structural variant models is illustrated in **FIG. 1**. First, protein sequence information is derived based on the genetic polymorphisms. The subject protein is then assigned to a protein superfamily in order to identify related proteins to be used as templates to construct a 3-D model of the protein. If the superfamily is not known, sequence analysis or structural similarity searches can be performed to identify related proteins for use as templates in homology modeling studies. Once the conserved regions of the model are assembled, *ab initio* loop prediction or *ab* A10 *initio* secondary structure generation techniques can be used to complete the model. Energetic refinement of the structure can be accomplished by performing molecular mechanics calculations, for example, using an ECEPP type forcefield or through molecular dynamics simulations, for example, using a modified AMBER type forcefield. If necessary, the structures can be dynamically refined, for example, by using a simulated annealing protocol (e.g.,

100 ps equilibration, 500 ps dynamics, up to 1000°K, 1 fs data collection). For quality control, the protein structural characteristics, for example, stereochemistry e.g., phi/psi and side chain angles), energetics (e.g., strain energy), packing profile (e.g., packing factor per residue) and hydrophobic packing are evaluated and required to meet acceptable criteria before the structures are used in further studies or input into a structural polymorphism database.

2. Creating 3-D structural polymorphism databases

After 3-D structural models are constructed for all protein structural variants, representing all known genetic polymorphisms, these can be inputted into a structural polymorphism relational database, along with associated structural or physical properties or clinical data (if available), as shown in **FIG. 1**. The databases can then be used to aid in structure-based drug design studies or for clinical analysis.

A10 Cont

The database is preferably interfaced to a molecular graphics package that includes 3-D visualization and structural analysis tools, to analyze similarities and variations in the protein structural variant models. (see, copending U.S. application Serial No. 09/272,814, filed March 19, 1999, which is incorporated by reference herein in its entirety). Briefly, U.S. application Serial No. 09/272,814 provides a database and interface for access to 3-D molecular structures and associated properties, which can be used to facilitate the design of potential new therapeutics. The interface also provides access to other structure-based drug discovery tools and to other databases, such as databases of chemical structures, including fine chemical or combinatorial libraries, for use in structure-focused high-throughput screening, as well as to a host of public domain databases and bioinformatics sites.

A relational database that collects multiple data files relating to the same molecular structure in the same subdirectory and that provides an interface to access all of the collected files from the same structure using the same user

interface program is also provided. The collected files include a variety of information and computer file formats, depending on the type of information to be conveyed to users of the database. In practice, a user communicates over a public network, such as the Internet, or over a controlled network, such as an internet, with a secure file server that controls access to the collected files, and the interface to the collected files is provided by a standard graphical user interface program that is widely available. In this way, a convenient means of searching molecular structure data for characteristics of interest is provided. Data searching, file viewing, and investigation of multiple representations of molecular structures from within a single viewing program can also be performed using the database and interface.

A10 cont

Please replace the paragraphs on page 18, lines 14-30, with the following paragraphs.

Data for a molecular structure are loaded into the database by specifying the file pathnames for the various data files that contain the different types of data, including the different molecule views. Using a browser to view the data files permits various helper applications, called plug-ins, to smoothly and transparently accept the different file formats and provide views to the user. The various data files of the database are organized in accordance with the database design when they are loaded into the database and are managed by a relational database management program.

A11

In addition to 3-D protein structures, as provided herein, the database can optionally contain associated biological or clinical data, such as drug resistance, side effects, efficacy, pharmacokinetics and other data, that correlate with or can be correlated with the structural variants. This information will be used for correlating observed clinical effects to specific structural variants and for predicting clinical responses and outcomes based on a patient's structural variants, i.e., genetic polymorphisms.

Please replace the paragraph on page 22, lines 9-17, with the following paragraph.

A12

The variants may also be used to track polymorphic variations in infectious organisms, such as viruses. For example, the human immunodeficiency viruses (HIVs) reverse transcriptase and protease have served as drug targets (see, Erickson *et al.* (1996) *Ann. Rev. Pharmacol. Toxicol 36*:545-571); their three-dimensional structures are known (see, *e.g.*, Nanni *et al.* (1993) *Perspectives in Drug Discovery and Design 1*:129-150; Kroeger *et al.* (1997) *Protein Eng. 10*:1379-1383). The clinical emergence of drug-resistant variants of these viruses has limited the long-term effectiveness of drugs targeted against these enzymes.

Please replace the paragraph on page 23, lines 15-22, with the following paragraph.

A13

In certain preferred embodiments, the free energy of binding of different drugs or potential drugs to each structural variant model can be calculated. The total free energy of binding is decomposed based on the interacting residues in the protein active site (see, e.g., Wang *et al.* (1996) *J. Am. Chem. Soc. 118*:995-1001; Wang *et al.* (1995) *J. Mol. Biol. 253*:473-492; Ortiz *et al.* (1995) *J. Med. Chem. 38*:2681-2691, which describes a computational method for deducing QSARs from ligand-macromolecule complexes).

Please replace the paragraph on page 27, lines 16-24, with the following paragraph.

A14

The methods provided herein, represent a further advance in the use of rational drug design methods. As described herein, shown herein, polymorphic variation has an effect upon the 3-D structure of encoded proteins. As a result, drugs interact with variants differently, leading to differential responses in the population as a whole. A new approach to drug design and testing is provided herein by identifying polymorphisms, and determining 3-D resulting structures,

A14 cont

which are then used in computation drug design or in selection of patient populations or in designing treatment protocols or other applications.

Please replace the paragraph on page 31, lines 1-8, with the following paragraph.

A15

The predicted correlations can also be used to aid in the design of subsequent clinical trials. The follow-up trials can be made more effective through the judicious selection of patients with given genotypes (i.e., those exhibiting the same genetic polymorphisms), as guided by the structurally predicted outcomes. For example, a clinical trial can be designed based on a subpopulation of clinical subjects which exhibit a specific genetic polymorphism (i.e. structural variant) to demonstrate the effectiveness of a given therapeutic on a targeted population.

Please replace the paragraph beginning on page 33, line 23, through page 34, line 13, with the following paragraph.

EXAMPLE 1

BINDING CORRELATIONS OF MUTANT FORMS OF HCV PROTEASE WITH DIFFERENT INHIBITORS

Introduction

A16

During HCV replication, the final steps of processing are performed by a virally encoded chymotrypsin-like serine protease NS3. NS3 is an approximately 3000 amino acid protein that contains, from the amino terminus to the carboxy terminus, a nucleocapsid protein (C), envelope proteins (E1 and E2) and several non-structural proteins (NS1, 2, 3, 4a, 4b, 5a and 5b). NS3 is an approximately 68 kda protein, encoded by approximately 1893 nucleotides of the HCV genome, and has two distinct domains: (a) a serine protease domain containing approximately 200 of the N-terminal amino acids; and (b) an RNA-dependent ATPase domain at the C-terminus of the protein. The NS3 protease is considered a member of the chymotrypsin family and is a serine protease that is responsible for proteolysis of the polypeptide (polyprotein) at the NS3/NS4a, NS4a/NS4b, NS4b/NS5a and NS5a/NS5b junctions responsible for generating

A16 cont

four viral proteins during viral replication. This protease is inhibited by N-terminal cleavage products of substrate peptides. The NS3 protease, which is necessary for polypeptide processing and viral replication has been identified, cloned and expressed (see, e.g., U.S. Patent No. 5,712,145).

Please replace the paragraph on page 36, lines 24-29, with the following paragraph.

Binding energies of the peptide-protein complexes

Binding energies were estimated using the equation:

$$E_{bind} = E_o + E_{compl} - E_{pept} - E_{prot},$$

A17

where E_{compl} is the energy of the complex, E_{pept} & E_{prot} are separate energies of the peptide and protein, respectively, and E_o is an adjustable constant.

Please replace the paragraph on page 38, lines 1-5, with the following paragraph.

Validation of the models: modifications of the protein and ligands in the binding site

A18

Mutation K136M and peptide modifications known from structure activity relationship (SAR) studies were performed in low-energy structures of the NS3-peptide 2 complex.

Please replace the paragraph beginning on page 39, line 24, through page 40, line 5, with the following paragraph.

EXAMPLE 2

LEAD OPTIMIZATION BY RECEPTOR-BASED FREE ENERGY QUANTIATIVE STRUCTURE ACTIVITY RELATIONSHIPS (QSARS) FOR TUMOR NECROSIS *factor* (TNF) RECEPTOR ANTAGONIST FINDING

A19

The goal of the modeling studies in this phase was to discover the binding modes and complex structures of the compounds that bind to TNF receptor type I protein, in order to guide design of new compounds. An approach that relies on docking compounds to the receptor, evaluating free energy changes of binding of the docked structures, and comparing the